Modeling the Spread of Influence for Independent Cascade Diffusion Process in Social Networks

Dr. Zesheng Chen and Kurtis Taylor

Indiana University – Purdue University Fort Wayne

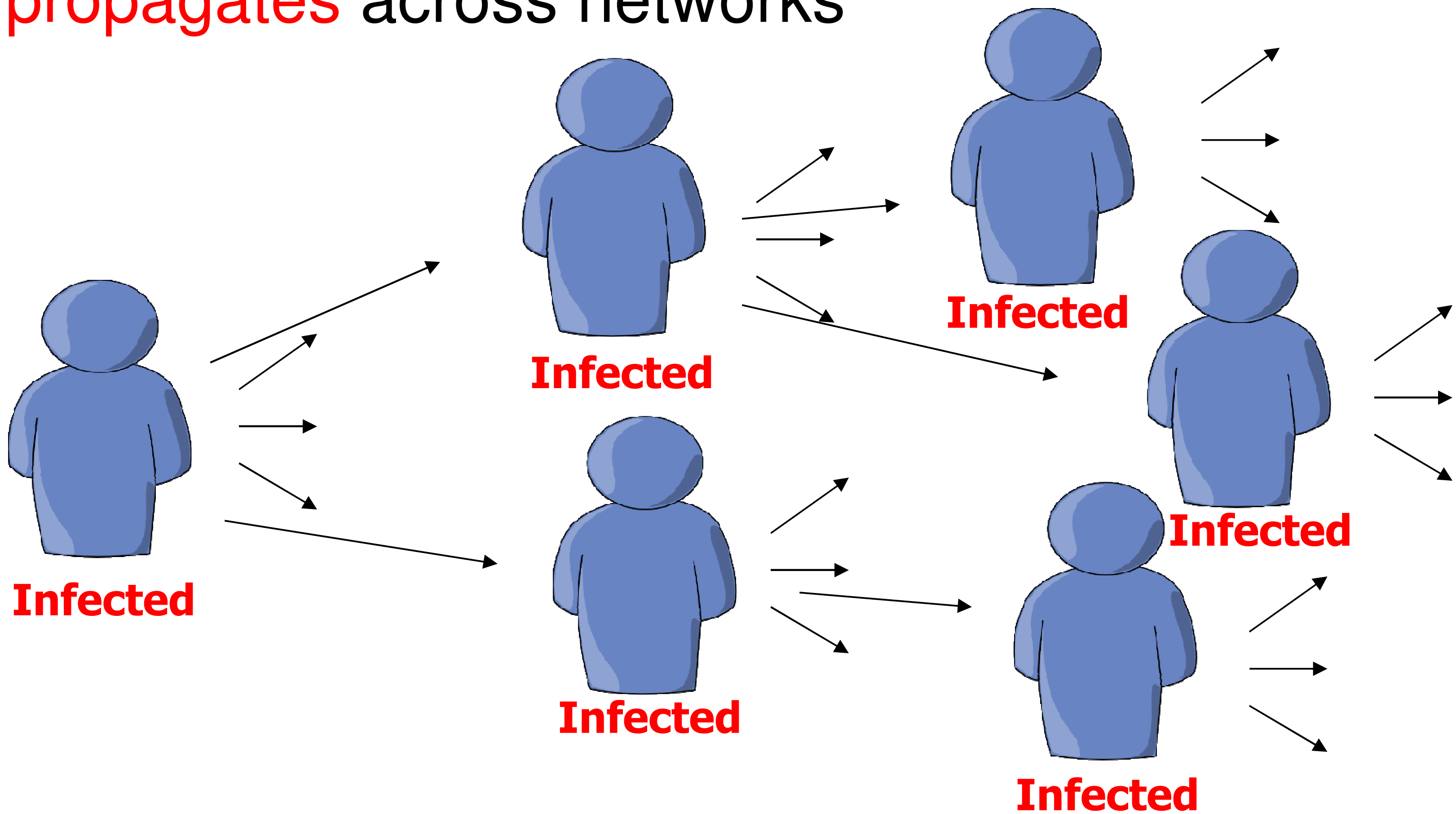
Social Influence in Online Social Networks

- Viral marketing (“word-of-mouth”)
- Blog information cascading
- Rumor spreading
- ……

Bear a resemblance to epidemic process!

Epidemic Process

- Epidemic process is a process that information self-propagates across networks



Classic Influence Diffusion Processes

- Two basic diffusion processes
  - Independent Cascade (IC)
  - Linear Threshold (LT)
- Study the *influence maximization problem*

In this work, we focus on the IC diffusion process.

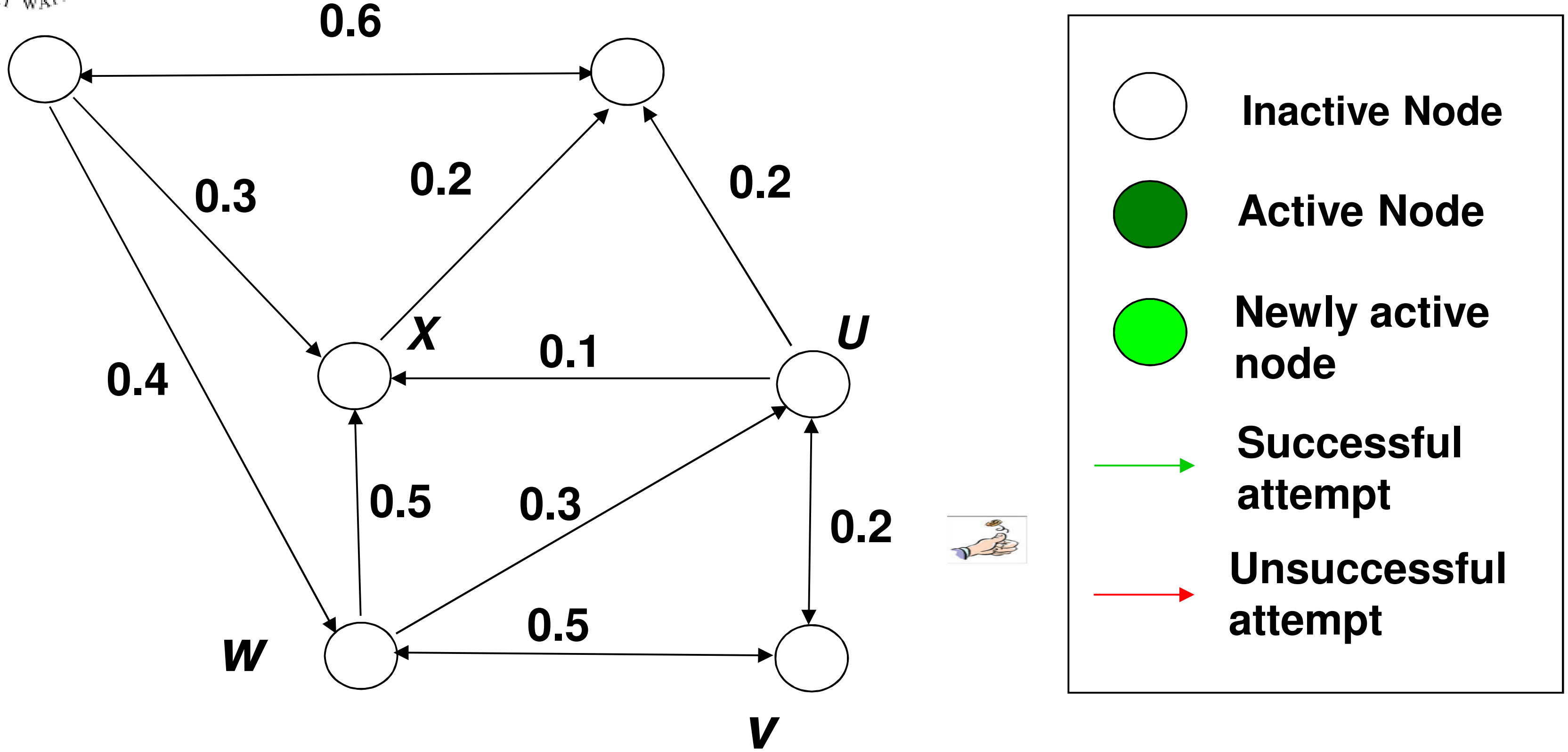
Refer to KDD'03 by Kempe, Kleinberg, and Tardos.

Independent Cascade (IC) Diffusion Process

- When a node in a social network becomes active (or infected), it has a single chance of activating (or infecting) each currently inactive neighbor.
- The activation attempt succeeds with a probability.

Refer to KDD'03 by Kempe, Kleinberg, and Tardos.

Example of IC Diffusion



Stop!

Refer to www.cs.cmu.edu/~xiaonanz/Maximizing-the-Spread-of-Influence.ppt

IC Diffusion Process

- Can be characterized by a susceptible-infected-recovered (SIR) mathematical model form epidemiology.

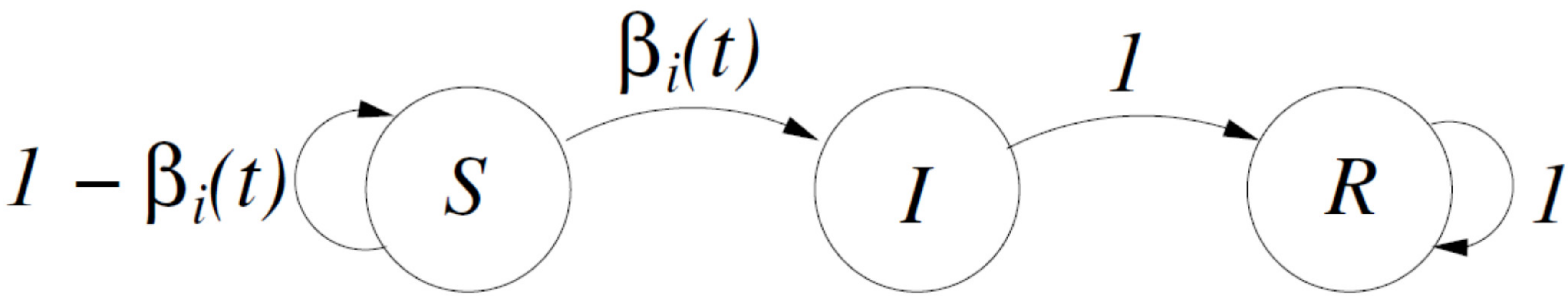


Fig. 1. SIR model for node i.

Questions

- How can we find an accurate mathematical model to characterize the spread of influence for the IC diffusion process in online social networks?
- Can spatial dependence among nodes affect the accuracy of mathematical models? If so, how significantly?
- How can such an accurate mathematical model help to solve the influence maximization problem?

Outline

- Mathematical Framework and Models
- Performance Evaluations

Mathematical Framework

- $X_i(t)$: status of node i at time t

$$X_i(t) = \begin{cases} \mathbf{0}, & \boldsymbol{if\ susceptible} \\ \mathbf{1}, & \boldsymbol{if\ infected} \\ \mathbf{-1}, & \boldsymbol{if\ recovered} \end{cases}$$

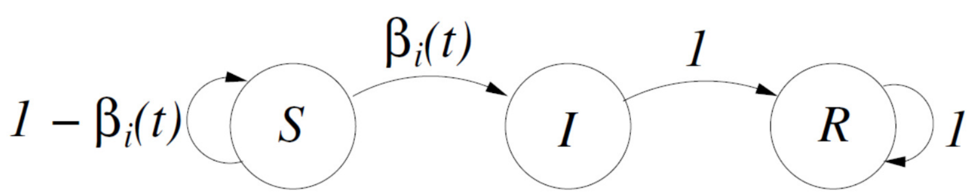


Fig. 1. SIR model for node i.

$$\boldsymbol{S}_i(t+\mathbf{1}) = \boldsymbol{S}_i(t)\,[\mathbf{1} - \boldsymbol{\beta}_i(t)]$$

$$\boldsymbol{I}_i(t+\mathbf{1}) = \boldsymbol{S}_i(t)\,\boldsymbol{\beta}_i(t)$$

$$\boldsymbol{R}_i(t+\mathbf{1}) = \boldsymbol{I}_i(t) + \boldsymbol{R}_i(t)$$

Mathematical Framework

$$\boldsymbol{\beta}_i(t) = \sum_{\boldsymbol{x}_{\boldsymbol{N_i}}(t)} \boxed{\boldsymbol{P}\left(\boldsymbol{X}_{\boldsymbol{N_i}}(t) = \boldsymbol{x}_{\boldsymbol{N_i}}(t) \middle| X_i(t) = 0\right)} \cdot \boldsymbol{f}_{\boldsymbol{i}}(t)$$

$$\boldsymbol{f}_i^{\boldsymbol{IC}}(t) = \mathbf{1} - \prod_{j \in \boldsymbol{N_i}} (\mathbf{1} - \beta_{ji})^{\frac{x_j^2(t) + x_j(t)}{\mathbf{2}}}$$

Independent Model

- Assume spatial independence between nodes

$$\boldsymbol{P}\left(\boldsymbol{X}_{\boldsymbol{N_i}}(t) = \boldsymbol{x}_{\boldsymbol{N_i}}(t) \middle| X_i(t) = 0\right) = \prod_{j \in \boldsymbol{N_i}} \boldsymbol{P}(X_j(t) = x_j(t))$$

$$\boldsymbol{\beta}_i^{IC\_ind}(t) = \mathbf{1} - \prod_{j \in \boldsymbol{N_i}} (\mathbf{1} - \boldsymbol{\beta}_{ji} I_j(t))$$

Such a model has been applied in previous work.

Markov Model

- Assume spatial Markov dependence
- Inspired by the local Markov property of Markov Random Field

$$\boldsymbol{P}\left(\boldsymbol{X}_{\boldsymbol{N}_{\boldsymbol{i}}}(t) = \boldsymbol{x}_{\boldsymbol{N}_{\boldsymbol{i}}}(t) \middle| X_i(t) = \boldsymbol{0}\right)$$

$$= \prod_{j \in \boldsymbol{N}_{\boldsymbol{i}}} \boldsymbol{P}(X_j(t) = x_j(t) | X_i(t) = \boldsymbol{0})$$

$$\boldsymbol{\beta}_i^{IC\_mar}(t) = \mathbf{1} - \prod\nolimits_{j \in \boldsymbol{N}_{\boldsymbol{i}}} [\mathbf{1} - \boldsymbol{\beta}_{ji} \boldsymbol{P}\left(X_j(t) = \mathbf{1} \mid X_i(t) = \mathbf{0}\right)]$$

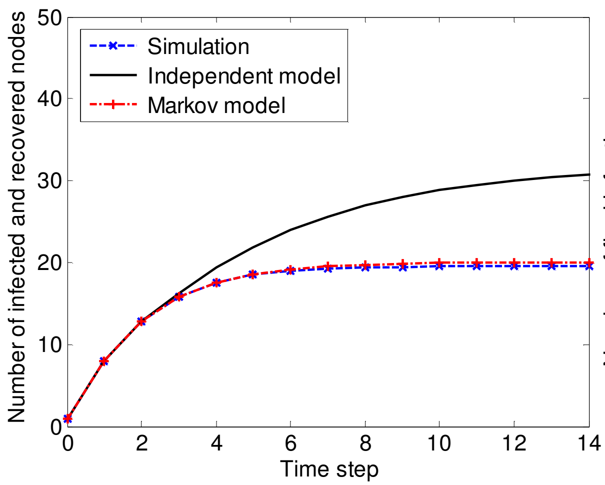
Outline

- Mathematical Framework and Models
- Performance Evaluations

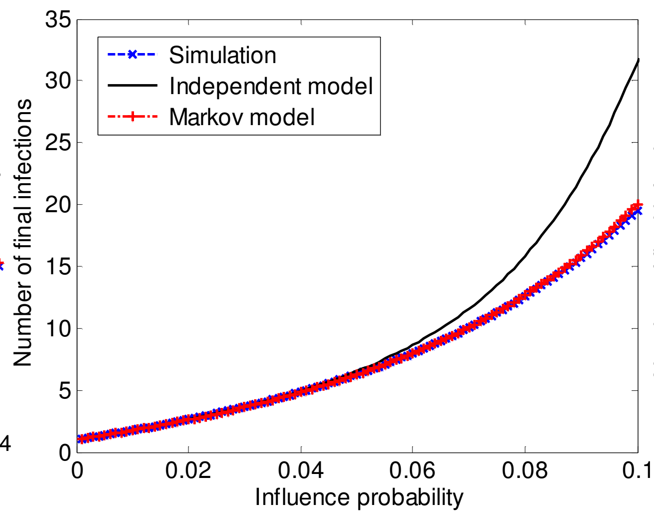
Simulation Setup

- Simulate the spread of influence for the IC diffusion process in an undirected graph
- Assume the influence probability is the same for all links and $0.001 \leq \beta \leq 0.1$
- Use discrete time and random number generator
- Run 20,000 times using different seeds for each scenario

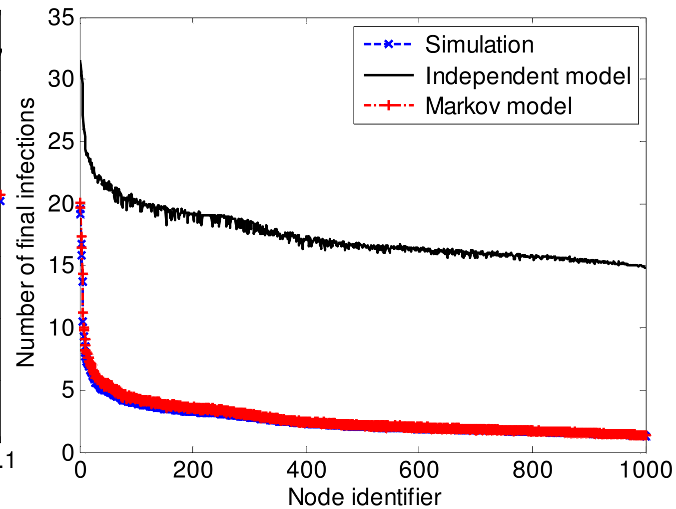
BA Power-Law Topology (1,000 Nodes)



β = 0.1

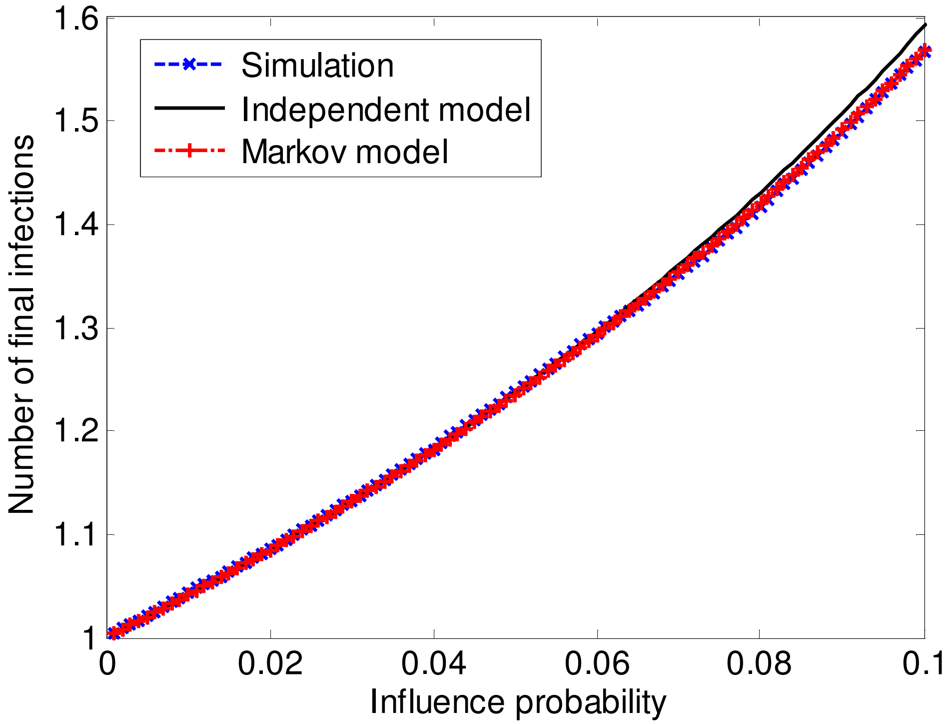


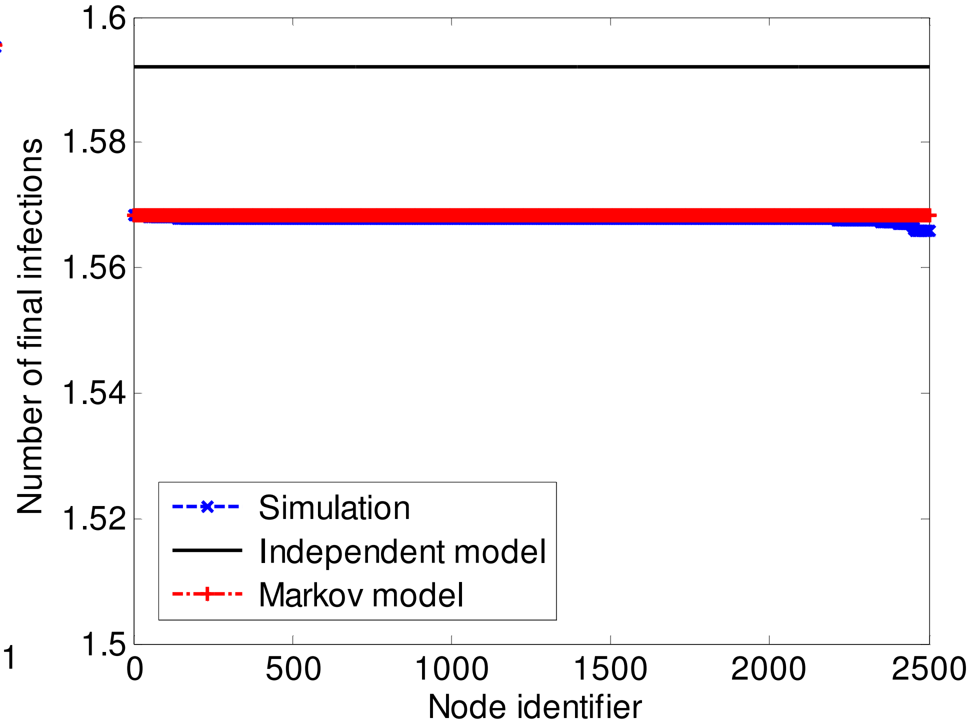
Influence of the node with the largest nodal degree



β = 0.1

Lattice Topology (2,500 Nodes)

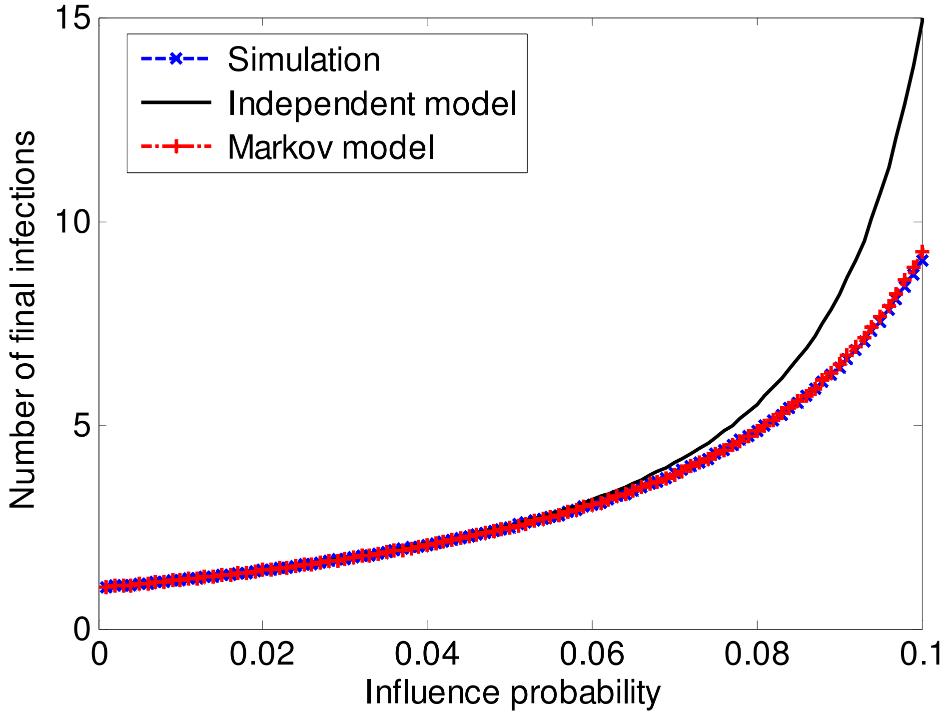


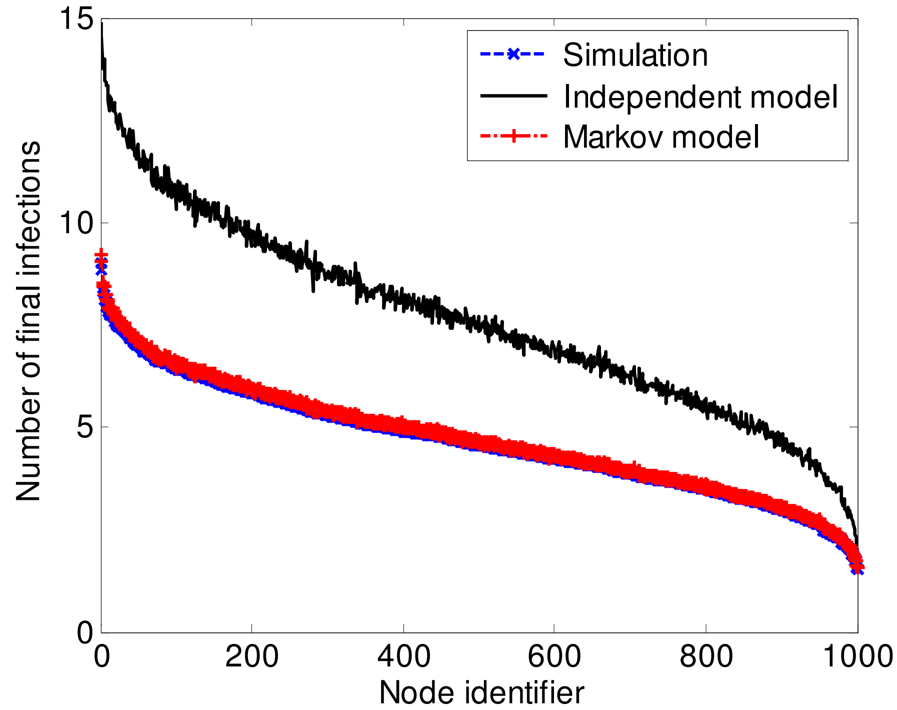


Influence of the node with the largest nodal degree

β = 0.1

ER Random Graph (1,000 Nodes)

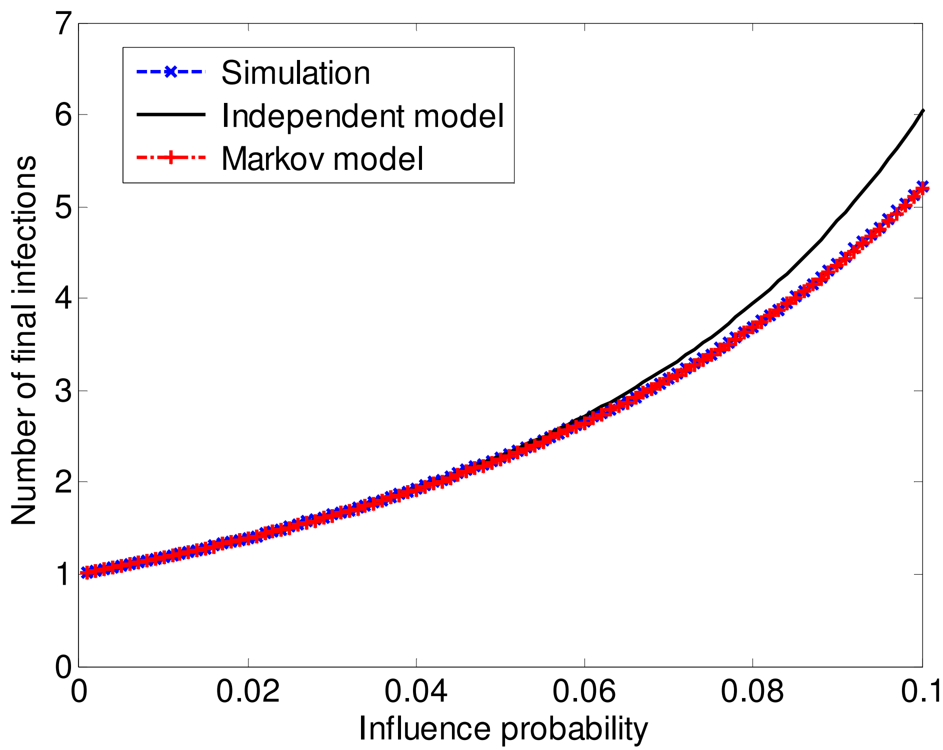




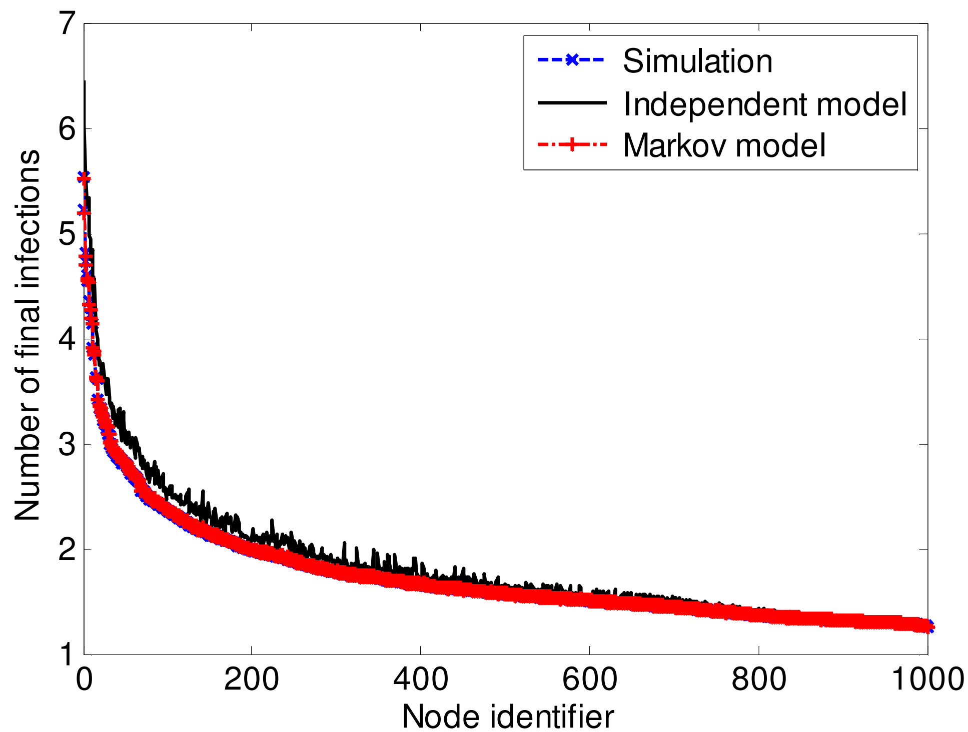
Influence of the node with the largest nodal degree

β = 0.1

Exponential Growth Random Graph (1,000 Nodes)



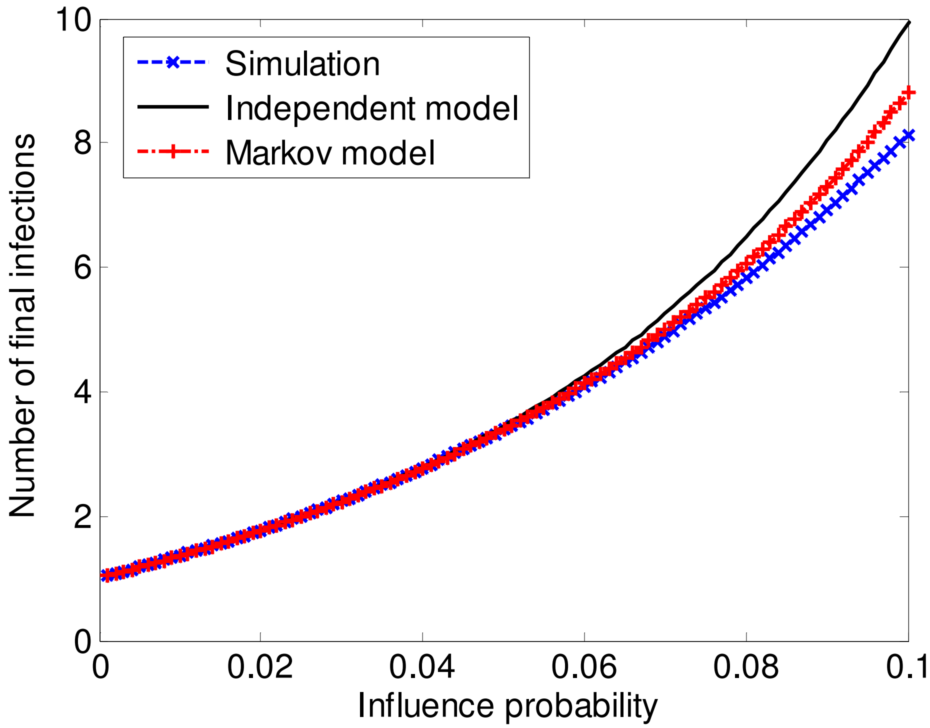
Influence of the node with the largest nodal degree



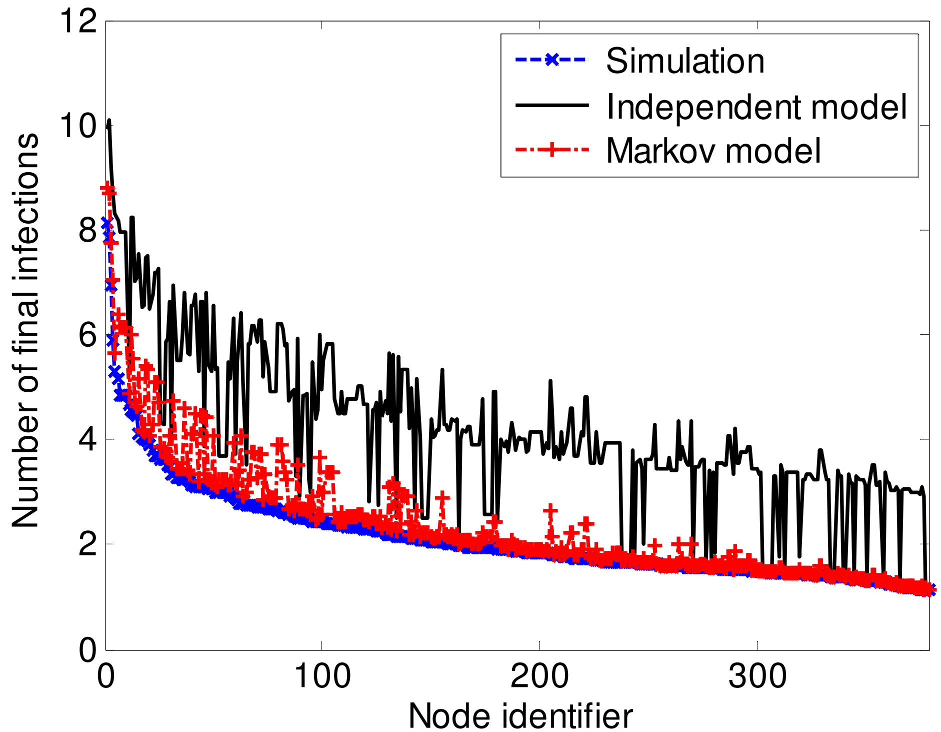
β = 0.1

A Real Topology

- Coauthorship network of scientists working on network theory and experiment.



Influence of the node with the largest nodal degree



β = 0.1

Time to Run Simulations and Markov Model

- Coauthorship network
  - Simulation took about 374 seconds
  - Markov model used only 6 seconds

Conclusions

- An accurate mathematical model needs to consider the spatial dependence among nodes in social networks.
- Spatial dependence among nodes significantly affect the accuracy of mathematical models.
  - Spatial Markov dependence
- Our Markov model can significantly reduce the time to predicate the influence of a node and can complement to the solutions to the influence maximization problem.

Thanks for your attention